SSC Statistical Society of Canada
Société Statistique du Canada

2000 Proceedings of the Survey Methods Section

Recueil 2000 de la Section des méthodes d'enquête

Papers presented at the 28th Annual Meeting of the Statistical Society of Canada, Ottawa, Ontario, Canada, June 4 - 7, 2000

Communications présentées au 28ième Congrès annuel de la Société statistique du Canada, Ottawa, Ontario, Canada du 4 au 7 juin 2000

SSC Annual Meeting, June 2000
Proceedings of the Survey Methods Section, pp. 185-190

A NONLINEAR HIERARCHICAL MODELLING APPROACH FOR CENSUS UNDERCOVERAGE ESTIMATION

Yong You[1]

ABSTRACT

Area-level nonlinear mixed effects models are considered in this paper for Canada census undercoverage estimation. We fit an area-level nonlinear mixed effects model to the province-level undercoverage survey estimates. In particular, the sampling model is based on the survey estimate of the undercoverage count, and the linking model is a log-linear model for the undercoverage rate. A full hierarchical Bayes (HB) approach is developed to obtain the posterior estimates of the census undercoverage using Markov Chain Monte Carlo (MCMC) sampling methods. Our result shows that the proposed method can provide efficient model-based estimates. Analysis of model fitting is also presented using posterior predictive distributions, and the corresponding result indicates that the proposed model fits the data quite well.

KEY WORDS: Census undercoverage, Gibbs sampling, Model checking, Nonlinear mixed model, Posterior.

RÉSUMÉ

Les modèles non linéaires à effets mixtes au niveau des régions sont considérés dans cet article dans le cadre de l'estimation de la sous-couverture présente au recensement canadien. Un modèle non linéaire à effets mixtes au niveau des régions est utilisé afin de modéliser les estimés de sous-couverture au niveau provincial. En particulier, le modèle d'échantillonnage est basé sur une estimation de la sous-couverture, et le modèle de liaison, servant à modéliser le taux de sous-couverture, est un modèle log-linéaire. Une approche hiérarchique de Bayes est développée afin d'obtenir des estimateurs *a posteriori* de la sous-couverture, en utilisant les méthodes de Monte Carlo avec chaîne de Markov (MCCM). Les résultats obtenus indiquent que la méthode proposée engendre des estimateurs, basés sur le modèle, qui sont efficaces. De plus, une évaluation du modèle, utilisant les distributions prédictives *a posteriori*, est présentée et les résultats indiquent que le modèle proposé semble adéquat.

MOTS CLÉS: sous-couverure au niveau du recensement, échantillonneur de Gibbs, évaluation du modèle, modèle non linéaire à effets mixtes, *a posteriori*.

1. INTRODUCTION

In Canada, a census is conducted every five years. However, the census does not enumerate all the inhabitants that should fill a census form on Census Day. In the 1991 Canadian census, it is estimated that about 3% of the population were not enumerated. Thus the census needs to be adjusted for undercoverage in order to properly represent the demographic picture of the country on Census Day. Since 1966, the Reverse Record Check (RRC) has been used by Statistics Canada to measure the gross number of persons missed by the census. Starting 1991, an Overcoverage Study was conducted to measure the gross number of persons erroneously included in the census. In 1991, for the first time, the RRC results were combined with those of the Overcoverage Study to produce the direct survey estimates of the net undercoverage for the nation and all provinces. Through the analysis of the results of these coverage studies, the collection methodology is adjusted in order to improve coverage in the succeeding census. More details on the coverage studies can be found, for example, in Germain and Julien (1993).

In 1991, the population estimates were based on the census counts adjusted for the estimated net undercoverage in the census. The base population was formed by adding the net provincial undercoverage estimate to the provincial census count. This created an adjusted base upon which all the other population figures were derived using modelling and demographic methods. Rivest (1995) proposed a composite estimator to estimate the provincial undercoverage using the national undercoverage rate as a synthetic estimate. The effect of the composite estimator is to shrink all provincial rates to the national rate. Rivest's composite estimator performs poorly at extreme provinces, namely, P.E.I. and Ontario, the

[1] Yong You, Household Survey Methods Division, Statistics Canada, Ottawa, Canada, K1A 0T6, yongyou@statcan.ca.

smallest and the largest provinces of Canada (You, 1997).

In recent years, modeling techniques have been applied to improve the direct undercoverage estimates from sample surveys. Datta *et al.* (1992) and Zaslavsky (1993) proposed hierarchical Bayes (HB) models for the adjustment of census undercoverage based on US census count and situations. Dick (1995) proposed a regression model utilizing an empirical Bayes (EB) method to estimate the undercoverage in small domains. Dick and You (1997) employed a linear Fay-Herriot model (Fay and Herriot, 1979) for census undercoverage rates in a HB framework. In this paper we are particularly interested in modeling the province-level census undercoverage counts. In particular, we propose a nonlinear mixed model in which the sampling model is based on the undercoverage counts whereas the linking model is based on rates. The undercoverage rate is expressed in terms of undercoverage counts. The proposed model overcomes some limitations of the model proposed by Dick and You (1997) and can provide model-based estimates for undercoverage counts and rates simultaneously. The remainder of the article is laid out as follows. In Section 2 we formally present the models for census undercoverage estimation. In Section 3 we illustrate the Gibbs sampling method based on the proposed nonlinear hierarchical models. In Section 4 we describe the implementation of the Gibbs sampler and present the estimation results. In Section 5 we describe diagnostics used to assess model fitting. Finally, we present some concluding remarks in Section 6.

2. UNDERCOVERAGE MODELS

Suppose there are m provinces. In the i-th province the census has counted c_i persons while an unknown number u_i persons are missed by the census. The coverage studies provide a survey estimate y_i of the net undercoverage along with an associated variance ξ_i^2. The true population of the i-th province can be written as $T_i = c_i + u_i$. Since the census count c_i is observed without sampling error, most of the work in constructing a model centres on the estimate of missed persons u_i. The coverage studies use sample surveys which through standard estimation procedures produce an estimate of the number of missed persons. The model generally used to describe this estimation situation can be written as

$$y_i = u_i + \varepsilon_i, \quad i = 1, \ldots, m, \qquad (1)$$

where $\varepsilon_i \sim N(0, \xi_i^2)$. This model assumes that the direct survey estimator y_i is design-unbiased. The design-unbiased assumption may be restrictive, in particular the estimates of missed persons y_i are probably subject to some unknown bias (Zaslavsky, 1993). However, since we don't have estimate for the possible bias as in the US case (Zaslavsky, 1993), we assume that the survey estimate y_i is design-unbiased for u_i. This is also a common assumption used in small area estimation (Rao, 1999). The normal distribution assumption of the sampling errors at the province level seems quite reasonable due to large sample size. The sampling variances ξ_i^2 are estimated through generalized variance function models of the form $V(y_i) \propto c_i^{\gamma}$ (Dick, 1995) and then treated as known in the model. Thus ξ_i^2 is actually a smoothed estimate of the sampling variance of y_i.

Instead of using (1) directly to model the undercoverage count u_i, Dick and You (1997) used a transformation and considered modelling undercoverage rate. Namely, let $\theta_i = u_i / (u_i + c_i)$ and $\hat{r}_i = y_i / (y_i + c_i)$, where θ_i is defined as the true undercoverage rate for the i-th province, and $\hat{r}_i$ is treated as the direct estimator of θ_i. Based on this transformation, Dick and You (1997) considered the following model:

$$\hat{r}_i = \theta_i + e_i, \quad \theta_i = x_i^T \beta + v_i, \quad i = 1, \ldots, m, \qquad (2)$$

where β is a vector of regression parameters; v_i is a normal random effect with $E(v_i) = 0$ and $V(v_i) = \sigma^2$, where σ^2 is unknown; $e_i \sim N(0, \psi_i^2)$. The sampling variance ψ_i^2 was treated as known and calculated from a Taylor approximation as $\psi_i^2 \approx (1-\theta_i)^4 V(y_i) / c_i^2 \approx (1-\hat{r}_i)^4 \xi_i^2 / c_i^2$. The model given by (2) is a standard application of Fay-Herriot model for small area estimation (Fay and Herriot, 1979). However, we note that model (2) has the following limitations: (i) the zero mean sampling model may not be true due to the nonlinear transformation. The results obtained from the model based on the transformed data could be significantly different from those obtained from the model based on the original survey estimates (You and Rao, 2000); (ii) the known sampling variance ψ_i^2 is a very strong assumption, since ψ_i^2 is indeed a function of the unknown mean θ_i; (iii) in model (2) assuming v_i to be normal may not be appropriate because θ_i is the undercoverage rate which is between 0 and 1, although similar models are also used by other people, for example, Datta et al. (1992) and Zaslavsky (1993). A more realistic model may use $\log(\theta_i) = x_i^T \beta + v_i$. To

overcome these limitations, we now propose a new model:

Model 1:

- Sampling model: $y_i = u_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \xi_i^2)$, where ξ_i^2 is known;
- Linking model: $\log(\frac{u_i}{u_i + c_i}) = x_i^T \beta + v_i$, $v_i \sim N(0, \sigma^2)$.

In Model 1, the sampling model is a linear model given directly by (1) and is based on the undercoverage count. The sampling variance is known. The linking model is a log-linear random effects model for the undercoverage rate, but here the rate θ_i is expressed as a function of count u_i, that is, $\theta_i = u_i / (u_i + c_i)$. In Model 1, we cannot combine the linking model with the sampling model to form a standard linear mixed effects model. Thus standard methods for linear mixed models cannot be applied. This kind of problems for small area estimation is discussed in Rao (1999) and You and Rao (2000). However, by using a complete HB approach with the Gibbs sampling method, we can find the posterior estimates of the undercoverage count and rate simultaneously.

3. GIBBS SAMPLER

We now use Model 1 to obtain model-based estimates of census undercoverage u_i. In a hierarchical Bayes frame, Model 1 is expressed as:

$$y_i \mid u_i \sim N(u_i, \xi_i^2), \quad i = 1, \ldots, m, \tag{3}$$

and

$$\frac{u_i}{u_i + c_i} \mid \beta, \sigma^2 \sim \log N(x_i^T \beta, \sigma^2), \quad i = 1, \ldots, m, \tag{4}$$

where logN denotes a log-normal density function. Priors of β and σ^2 are set as: $\pi(\beta) \propto 1$ and $\sigma^2 \sim IG(a, b)$ with a and b known positive constants. Let $Y = (y_1, \ldots, y_m)^T$ and $U = (u_1, \ldots, u_m)^T$, we are interested in the posterior distribution of u_i given the data Y, in particular, posterior mean $E(u_i \mid Y)$ and posterior variance $V(u_i \mid Y)$. The Gibbs sampling method (Gelfand and Smith, 1990) is used to simulate samples for u_i. To implement the Gibbs sampling, we need to draw samples from the following full conditional distributions:

- $[u_i \mid Y, \beta, \sigma^2] \propto \frac{u_i + c_i}{u_i} \exp\{-\frac{(y_i - u_i)^2}{2\xi_i^2} - \frac{(\log(u_i/(u_i + c_i)) - x_i^T \beta)^2}{2\sigma^2}\}$, $i = 1, \ldots, m$;
- $[\beta \mid Y, U, \sigma^2] \sim N((\sum_{i=1}^m x_i x_i^T)^{-1}(\sum_{i=1}^m x_i \log\frac{u_i + c_i}{u_i}), \sigma^2(\sum_{i=1}^m x_i x_i^T)^{-1})$;
- $[\sigma^2 \mid Y, U, \beta] \sim IG(a + m/2, b + \sum_{i=1}^m (\log\frac{u_i + c_i}{u_i} - x_i^T \beta)^2 / 2)$.

Drawing samples from $[\beta \mid Y, U, \sigma^2]$ and $[\sigma^2 \mid Y, U, \beta]$ is straightforward. However, $[u_i \mid Y, \beta, \sigma^2]$ does not have a closed form. To draw samples from $[u_i \mid Y, \beta, \sigma^2]$, Metropolis-Hastings updating scheme (see, e.g., Chib and Greenberg, 1995) is used within the Gibbs sampler. The Metropolis-Hastings updating step is summarized as follows: Suppose the Markov chain is at the k-th iteration. To update $[u_i \mid Y, \beta, \sigma^2]$, we first draw a candidate sample $u_i^{(k+1)}$ from $N(y_i, \xi_i^2)$, then with probability

$$\alpha(u_i^{(k)}, u_i^{(k+1)}) = \min\{\frac{g(u_i^{(k+1)}, \beta^{(k)}, \sigma^{2(k)})}{g(u_i^{(k)}, \beta^{(k)}, \sigma^{2(k)})}, 1\}, \tag{5}$$

we accept this $u_i^{(k+1)}$; otherwise, set $u_i^{(k+1)} = u_i^{(k)}$. In (6), $g(\cdot)$ is a function of u_i, β, σ^2 given by

$$g(u_i, \beta, \sigma^2) = \frac{u_i + c_i}{u_i} \exp\{-\frac{(\log(u_i/(u_i + c_i)) - x_i^T \beta)^2}{2\sigma^2}\}. \tag{6}$$

The estimation of the posterior distribution of u_i as well as $E(u_i \mid Y)$ and $V(u_i \mid Y)$ can be based on the samples $\{u_i^{(k)}\}$ from the Gibbs sampler.

4. IMPLEMENTATION AND ESTIMATION

As in Dick and You (1997), we used 1991 data in our analysis. We used log-transformation of the census count as the auxiliary variable, that is, $x_{1i} = \log(c_i)$, then the linking model for u_i is $\log(u_i/(u_i + c_i)) = \beta_0 + x_{1i}\beta_1 + v_i$. To implement and monitor the convergence of the Gibbs sampler, we follow the basic approach given in Gelman and Rubin (1992). We independently simulated L=8 sequences, each of length t=2d, with d=5000. The first 5000 iterations of each sequence were deleted. To reduce the autocorrelation in the sequence, we took every 10th iteration for the left 5000 iterations, leading to 500 iterations for each sequence kept for analysis. Thus we finally have L=8 sequences with size n=500 for each sequence.

For the parameter of interest u_i, let u_{ilj}^* denote the j-th simulated value of u_i in the l-th sequence. Then the

posterior mean of u_i is estimated by $\hat{u}_i = \sum_{l=1}^{L}\sum_{j=1}^{n} u^*_{ilj} / Ln$. Then we computed B_i / n, the variance between the 8 sequence means, $\bar{u}_{il.}$ each based on n=500 simulated values of u_i; that is, $B_i / n = \sum_{l=1}^{L}(\bar{u}_{il.} - \hat{u}_i)^2 / (L-1)$. Also, let W_i denote the average of the 8 within-sequence variance, s^2_{il}, each based on n-1 degrees of freedom; that is, $W_i = \sum_{l=1}^{L} s^2_{il} / L$. Then the posterior variance of u_i is estimated by a weighted average of W_i and B_i, namely, $\hat{\sigma}^2_i = (n-1)W_i / n + B_i / n$, where n=500. Note that if only one sequence is simulated, B_i cannot be calculated. To estimate the i-th province undercoverage rate θ_i, which is defined as $\theta_i = u_i / (u_i + c_i)$, let $\theta^*_{ilj} = u^*_{ilj} / (u^*_{ilj} + c_i)$, then the posterior mean of θ_i can be estimated by $\hat{\theta}_i = \sum_{l=1}^{L}\sum_{j=1}^{n} \theta^*_{ilj} / Ln$. Similarly, we can calculate the posterior variance of θ_i. To monitor the convergence of the Gibbs sampler, we calculate $\hat{V}_i = \hat{\sigma}^2_i + B_i / Ln$, and then find $\hat{R}_i = \hat{V}_i / W_i$ for each observation. $\hat{R}_i$ is known as a potential scale reduction factor (Gelman and Rubin, 1992). If $\hat{R}_i$'s are near 1 for all of the parameters u_i of interest, then this suggests that the desired convergence is achieved in the Gibbs sampler. In our study, values of $\hat{R}_i$'s for the 10 provinces are equal to or very close to 1, they strongly suggest that the Gibbs sampler converged very well. Table 1 shows the direct undercoverage estimates and the posterior undercoverage estimates as well as the associated coefficients of variation. The posterior estimate of the census undercoverage count u_i has smaller coefficient of variation (CV) than the direct survey estimate y_i for all provinces except Ontario, Quebec and New Brunswick. Particularly for some small provinces such as P.E.I. and Manitoba, the improvement of posterior estimate over the direct estimate is significant. For Ontario and Quebec, the largest two provinces in Canada, the direct estimate and the posterior estimate have equal CV due to the large sample size for these two provinces. An interesting result is that for New Brunswick, the posterior estimate $\hat{u}_i$ has larger CV than the direct estimate y_i, and also $\hat{u}_i$ is quite different from y_i. This is more obvious in terms of undercoverage rate, where directly estimated rate $\hat{r}_i$ is 3.25% and the posterior rate $\hat{\theta}_i$ is only 2.55% for New Brunswick. Actually province New Brunswick is an outlier in the model. More detailed analysis will be given in the next section.

5. TEST OF MODEL FITTING

To test the overall fit of Model 2, we use the method of posterior predictive p value (Meng, 1994; Gelman *et al.*, 1995). Let $T(y,\theta)$ be a discrepancy measure depending on the data y and on parameters θ. Let θ^* represent a draw from the posterior distribution of θ and let y^* represent a draw from $f(y|\theta^*)$, then marginally y^* is a sample from the posterior predictive distribution $f(y | y_{obs})$, where y_{obs} represents the observed data. The posterior predictive p value is defined as $p = \Pr(T(y^*,\theta) > T(y_{obs},\theta) | y_{obs})$. Note that the probability is with respect to the posterior distribution given the observed data.

This is a natural extension of the usual p value in a Bayesian context. If a model fits the observed data, then the two values of the discrepancy measure are similar. In other words, if the given model adequately fits the observed model, then $T(y_{obs},\theta)$ should be

Table 1. Census undercoverage estimation

Province	y_i	$CV(y_i)$	$\hat{u}_i$	$CV(\hat{u}_i)$	$\hat{r}_i$%	$CV(\hat{r}_i)$	$\hat{\theta}_i$%	$CV(\hat{\theta}_i)$
NFLD	11566	0.16	10782	0.14	1.99	0.16	1.86	0.13
PEI	1220	0.30	1486	0.19	0.93	0.30	1.13	0.19
NS	17329	0.20	17412	0.14	1.89	0.20	1.90	0.14
NB	24280	0.14	18948	0.17	3.25	0.13	2.55	0.17
QUE	184473	0.08	189599	0.08	2.58	0.08	2.65	0.08
ONT	381104	0.08	368424	0.08	3.64	0.08	3.52	0.08
MAN	20691	0.21	21504	0.14	1.86	0.20	1.93	0.14
SASK	18106	0.19	18822	0.14	1.80	0.18	1.87	0.13
ALTA	51825	0.15	55392	0.12	2.01	0.14	2.13	0.12
BC	92236	0.10	89929	0.09	2.73	0.10	2.67	0.09

near the central part of the histogram of the $T(y^*,\theta)$ values if y^* is generated repeatedly from the posterior predictive distribution. Consequently, the posterior predictive p value is expected to be near 0.5 if the model adequately fits the data. Extreme p values (near 0 or 1) suggest poor fit. Computing the p value is relatively easy using the posterior simulation from the Gibbs sampler. For each simulated value θ^*, we can simulate y^* from the model and compute $T(y^*,\theta^*)$ and $T(y_{obs},\theta^*)$. Then the p value is estimated by the proportion of times that $T(y^*,\theta^*)$ exceeds $T(y_{obs},\theta^*)$. In the present context of census undercoverage, the discrepancy measure that we used for overall fit is $T(y,u)=\sum_i (y_i-u_i)^2/\xi_i^2$. Similar discrepancy measure is also used, for example, in Datta *et al.* (1999). We computed the p values for all 8 parallel runs. Under Model 1, the smallest p value is 0.374, the largest one is 0.482, and the average p value is 0.433. Thus we have no indication of any lack of overall fit. The p value strongly suggests the adequacy of the model.

We also computed two statistics that are useful in assessing model fit at the individual observation level. First, we computed $p_i^* = \Pr(y_i^* < y_{i(obs)} \mid y_{obs})$, which provides information on the degree of consistent overestimation or underestimation of the observed data. Second, we computed $r_i^* = E^*(y_i - y_{i(obs)} \mid y_{obs})/\sqrt{Var(y_i \mid y_{obs})}$, the standardized predictive residual. So we used the mean of the predictive distribution for y_i as the center and the variance as a measure of the spread. These standardized residuals can be used to compare model fit for individual observations. We used the posterior simulations y_i^* to compute p_i^* and r_i^*. Under Model 2, p_i^* ranges from 0.28 to 0.87, with a mean of 0.51 and a median of 0.49, indicating no consistent overestimation or underestimation. Among the 10 provinces, PEI has p_i^*=0.28 and New Brunswick has p_i^*=0.87, indicating that the model may not be appropriate for these two provinces, especially for New Brunswick. For the standardized residuals r_i^*, New Brunswick has the largest positive r_i^*=1.13, which indicates that the current model underestimates the census undercoverage count with respect to the observed value. PEI has the largest negative r_i^*=-0.57, which indicates that the model overestimates the undercount compared to the observed value. In terms of absolute value of r_i^*, New Brunswick has the largest standardized residual. p_i^* and r_i^* can also be used as criteria for outlier detection. An outlier is defined as an observation that is extreme (in the value of some statistic) relative to its predictive distribution under the model. Because outliers represent some data patterns that are not well described by the model, they suggest areas in which modifications of the model should be considered. In terms of the values of p_i^* and r_i^*, New Brunswick is obviously an outlier. Dick and You (1997) also obtained similar findings. Since New Brunswick always appears to be an outlier in various models, we suggest that further work should be done for the direct survey estimate of the census undercoverage for New Brunswick.

6. CONCLUSION REMARKS

In this paper, we have proposed a nonlinear mixed model for the problem of census undercoverage estimation. The model involves a linear sampling model and a log-linear linking model. We cannot combine the sampling model with the linking model to form a usual linear mixed model to which standard methods can be applied. However, with the Gibbs sampling method and Metropolis-Hastings algorithm, we find the HB estimator of the census undercoverage numerically. Our data analysis shows that the proposed model fit the data quite well. The HB method improves the direct survey estimates and also can be used to identify possible outliers in the data. This paper itself also provides a good example for the case of unmatched sampling and linking models in small area estimation as discussed in You and Rao (2000).

ACKNOWLEDGEMENT

The author would like to thank Dr. J.N.K. Rao of Carleton University and Dr. Mayer Alvo of University of Ottawa for helpful discussion.

REFERENCES

Chip, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49, 327-335.

Datta, G.S., Ghosh, M., Huang, E.T., Isaki, C.T., Schultz, L.K. and Tsay, J.H. (1992) Hierarchical and empirical Bayes methods for adjustment of census undercount: The 1988 Missouri dress rehearsal data. *Survey Methodology*, 18, 95-108.

Datta, G.S., Lahiri, P., Maiti, T. and Lu, K.L. (1999) Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association*, 94, 1074-1082.

Dick, P. (1995) Modelling net undercoverage in the 1991 Canadian census. *Survey Methodology*, 21, 45-54.

Dick, P. and You, Y. (1997) A hierarchical Bayes analysis of Census undercoverage. *Proceedings of Symposium 97 New Directions in Surveys and Censuses*, 101-105, Statistics Canada, Ottawa.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Gelman, A. and Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.

Germain, M.F., and Julien, C. (1993) Results of the 1991 census coverage error measurement program. *Proceedings of Seventh Annual Research Conference*, United States Bureau of the Census, 55-70. Washington.

Meng, X.L. (1994) Posterior predictive p value. *The Annals of Statistics*, 22, 1142-1160.

Rao, J.N.K. (1999) Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175-186.

Rivest, L.P. (1995) A composite estimator for provincial undercoverage in the Canadian census. *Proceedings of the Survey Method Section*, Statistical Society of Canada Annual Meeting, Montreal, Quebec, July, 1995.

You, Y. (1997) Hierarchical Bayes models for census undercoverage estimation. Internal Statistics Canada Report.

You, Y. (1999). *Hierarchical Bayes and Related Methods for Model-Based Small Area Estimation*. Ph.D. Thesis, Carleton University, Ottawa, Canada.

You, Y. and Rao, J.N.K. (2000) Small area estimation with unmatched sampling and linking models. SSC Annual meeting, Ottawa.

Zaslavsky, A.M. (1993) Combining census, dual-system, and evaluation study data to estimate population shares. *Journal of the American Statistical Association*, 88, 1092-1105.

C-1